

**Running Title: Circular Dichroism of Biomolecules**

**CIRCULAR DICHROISM AND ITS APPLICATION TO THE STUDY  
OF BIOMOLECULES**

**Stephen R. Martin<sup>1\*</sup> and Maria J. Schilstra<sup>2</sup>**

**<sup>1/</sup> Division of Physical Biochemistry**

**National Institute for Medical research**

**The Ridgeway, Mill Hill, London NW7 1AA, U.K.**

**<sup>2/</sup> Biological and Neural Computation Group**

**Science & Technology Research Institute**

**University of Hertfordshire, College Lane**

**Hatfield AL10 9AB, U.K.**

**(\*) To whom correspondence should be addressed.**

**e-mail: [s-martin@nimr.mrc.ac.uk](mailto:s-martin@nimr.mrc.ac.uk)**

## I. Introduction

Circular dichroism (CD), the differential absorption of left- and right-handed circularly polarised light, is a spectroscopic property uniquely sensitive to the conformation of molecules, and so has been very widely used in the study of biomolecules. CD often provides important information about the function and conformation of biomolecules that is not directly available from more conventional spectroscopic techniques such as fluorescence and absorbance. The experimentally measured parameter in CD is the difference in absorbance for left- and right-handed circularly polarised light,  $\Delta A (= A_L - A_R)$ . Because CD is an absorption phenomenon, the chromophores that contribute to the CD spectrum are exactly the same as those contributing to a conventional absorption spectrum. In order to show a CD signal a chromophore must be either inherently chiral (asymmetric) or must be located in an asymmetric environment. Chromophores that contribute to protein CD (see below) are generally achiral because they have a plane of symmetry. It is the interaction between the chromophores in the chiral field of the protein that introduces the perturbations leading to optical activity (Sreerama and Woody, 2004a). The near-UV CD bands of proteins (310 - 255 nm) derive from Trp, Tyr, Phe, and cystine (Note: cysteine does not absorb in this region) and reflect the tertiary, and occasionally quaternary, structure of the protein. Although several amino acid side chains (notably Tyr, Trp, Phe, His, and Met) absorb light strongly in the far-UV region of the spectrum (below 250 nm) the most important contributor here is the peptide bond (amide chromophore), with  $n \rightarrow \pi^*$  and  $\pi \rightarrow \pi^*$  transitions at  $\sim 220$  nm and  $\sim 190$  nm, respectively. The far-UV CD bands of proteins reflect the secondary structure of the protein ( $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn, and unordered content). In the case of nucleic acids and oligonucleotides the aromatic bases are the principal chromophores, with absorption beginning at around 300 nm and extending far into the vacuum UV region. The electronic

transitions of the ether and hydroxyl groups of the sugars begin at 200 nm, but their intensity is much weaker than that of the bases, and the electronic transitions of the phosphate groups begin further still into the vacuum UV (Johnson, 1996a).

Although CD spectroscopy generally provides only low-resolution structural information it does have two major advantages. First, it is extremely sensitive to changes in conformation, whatever their origin, and second, an extremely wide range of solvent conditions is accessible to study with relatively small amounts of material. The principal applications of CD spectroscopy in the study of biomolecules are:

- (a) The estimation of protein secondary structure content from far-UV CD spectra.
- (b) The detection of conformational changes in proteins and nucleic acids brought about by changes in pH, salt concentration, and added co-solvents (simple alcohols, tri-fluoroethanol etc.), and the structural analysis of recombinant native proteins and their mutants.
- (c) Monitoring protein or nucleic acid unfolding brought about by changes in temperature or by the addition of chemical denaturants (such as urea and guanidine hydrochloride).
- (d) Monitoring protein-ligand, protein-nucleic acid, and protein-protein interactions.
- (e) Studying (in favourable cases) the kinetics of macromolecule-macromolecule macromolecule-ligand interactions (particularly slow dissociation processes), and the kinetics of protein folding reactions. The general principles of the common kinetic methods are discussed in Eccleston et al. (This Volume) and will not be considered in detail here.

There are numerous excellent reviews that describe the basic principles of CD spectroscopy and its applications in the study of different biomolecules (Gray, 1996; Greenfield, 1996, 2004; Johnson 1985, 1988, 1990, 1996a; Strickland, 1974; Venyaminov and Yang, 1996; Woody 1985, 1995, 1996; Woody and Dunker, 1996; Yang et al., 1986).

## **II. Instrumentation and sample preparation**

### **A. Instrumentation**

CD instruments are commercially available from several sources: Jasco Inc. (<http://www.jascoinc.com/>), Aviv Biomedical Inc. (<http://www.avivbiomedical.com/>), OLIS Inc. (<http://www.olisweb.com/>), and Applied Photophysics (<http://www.photophysics.com/>). A Peltier system for temperature control and thermal ramping is an invaluable accessory, particularly for studies of the thermal unfolding of proteins and nucleic acids. The only other significant requirement is for a set of high quality quartz cuvettes with good far-UV transmission (either rectangular or cylindrical) with path lengths ranging from 0.1 to 10 mm. Self-masking (black-walled) micro or semi-micro cuvettes with 10 mm path length are particularly useful for near-UV CD measurements with small volumes (~ 0.25 ml). Cuvettes are obtainable from several suppliers (e.g., Hellma; <http://www.hellma-worldwide.de/>). Cuvettes with path lengths of less than 1 mm should always be calibrated. This is easily done by using the cuvette to record a conventional absorption spectrum of any solution with accurately known absorbance, or as described by Johnson (1996b). Cuvettes may have some intrinsic strain that can give significant CD artifacts and although moderate strain can be tolerated it is probably sensible to eliminate any possible strain effects by always orienting the cuvette the same way in the instrument. Cuvettes should always be cleaned immediately after use in order to avoid the build up of hard to remove protein deposits. This can be done using a preparation such as Hellmanex II cuvette cleaning solution. After cleaning, the cuvettes should be rinsed extensively with distilled water, then ethanol, and dried using an air pump or by evaporation. Cuvettes should be stored in the cases generally provided by the manufactures. All standard reagents used should be of the highest purity available. It is

particularly important that any organic solvents used should be of spectroscopic and should be checked for the absence of absorbing impurities.

## **B. Instrument care and calibration**

The CD instrument should always be purged with high-purity, oxygen-free, nitrogen (generally run at  $\sim 3$ -5 liters/min) for at least 20 minutes before starting the light source and throughout the measurements. If oxygen is present it may be converted to ozone by the far-UV light from the high intensity arc, and ozone will damage the expensive optical surfaces. Higher nitrogen flow rates will generally be necessary for measurements made at very short wavelengths.

The calibration of the instrument should be checked periodically. Although several CD standards are available, the one used most frequently is d10 camphor sulfonic acid (d10-CSA). The exact concentration ( $C$ ) of a solution of d10-CSA in water (at  $\sim 2.5$  mM) should be determined from an absorption spectrum (using  $\epsilon_{285} = 34.5 \text{ M}^{-1}\text{cm}^{-1}$ ) and not by weight because the solid is hygroscopic. The differential absorption ( $\Delta A$ ) recorded at 290.5 nm in a 10 mm path length cuvette should be  $C \cdot \Delta \epsilon_{M,290.5}$  (or  $32980 \cdot C \cdot \Delta \epsilon_{M,290.5}$  millidegrees - see section IV.A.), where  $\Delta \epsilon_{M,290.5} = 2.36 \text{ M}^{-1}\text{cm}^{-1}$  (Johnson, 1990). If the intensity is not within 1% of the expected value the user should refer to the manufacturer's handbook for details of the appropriate adjustment procedure. It is also advisable to check the wavelength calibration of the instrument and its general transmission performance in the far-UV from time to time. Because d10-CSA has a second CD band at 192.5 nm ( $\Delta \epsilon_{M,192.5} = -4.72 \text{ M}^{-1}\text{cm}^{-1}$ ) one can check the far-UV performance by recording the spectrum of the standard d10-CSA solution using a 1 mm path length cuvette. If the (absolute) value of the ratio of the intensities of these peaks is significantly less than 1.95 then the machine is no longer performing correctly; this is probably due to the age of the lamp and/or degradation of the mirrors, most probably the first

mirror after the light source. This measurement also provides a useful check on the wavelength calibration of the instrument although this can also simply be done by scanning with a holmium oxide filter in the light path and monitoring the voltage on the instrument's photomultiplier.

### **C. Sample Preparation**

All samples should, of course, be of the highest possible purity. Misleading results can be obtained even with relatively low levels of impurities if these have strong CD signals. For example, the weak near-UV signals of proteins can be swamped by the strong signals from relatively small levels of contaminating nucleic acids. One major problem in CD measurements is that the signals become seriously distorted if too little light reaches the photomultiplier and, in practical terms, this means that one cannot make reliable measurements on samples with an absorbance (sample plus solvent) much greater than about 1. The absorption spectrum of the sample should always be checked to see if and where this absorbance limit is going to be exceeded. In far-UV measurements the absorbance of the sample itself is generally rather small and the major problems arise from absorption by buffer components, almost all of which will limit far-UV penetration to some extent. The majority of simple buffer components will generally permit CD measurements to below 200 nm (see Johnson, 1996b). However, high concentrations of chloride and (especially) nitrate (use perchlorate or fluoride salts where possible), certain solvents (dioxane, DMSO), high concentrations (> 25 mM) of some biological buffers (Hepes, Pipes, Mes), high concentrations (> 0.25 mM) of common chelators (EGTA/EDTA), and high concentrations (> 1mM) of reducing agents (dithiothreitol and 2-mercaptoethanol) should be avoided whenever possible. It is also worth noting that distilled water stored in a polyethylene bottle will generally develop poor far-UV transparency owing to the presence of eluted polymer

additives. The CD spectra of membrane proteins are often recorded in detergent solubilised form in order to avoid artifacts arising from differential light scattering and absorption flattening (Fasman, 1996).

The far-UV CD spectra of proteins (260-178 nm) are intense and relatively small amounts of material are required to record them. Because all peptide bonds contribute to the observed spectrum the amount of material required (measured in mgs/ml) is effectively the same for any protein. Measurements are almost invariably made in short path length cuvettes in order to reduce absorption by buffer components (see above). Typical quantities are 200  $\mu$ l of a 0.1 - 0.15 mg/ml solution when using 1 mm path length cuvette or 30  $\mu$ l of a 1.0 - 1.5 mg/ml solution when using a 0.1 mm (demountable) cuvette. The latter is preferable for good far-UV penetration but the material is not generally recoverable.

The near-UV CD spectra of proteins are generally more than an order of magnitude weaker than the far-UV CD spectra. Recording them therefore requires more concentrated material and/or longer optical path lengths. Spectra are usually recorded under conditions similar to those used for measuring a conventional absorption spectrum, e.g., use a 10 mm path length cuvette and aim for a peak absorbance in the range 0.7-1.0 (The optimal absorbance for best signal to noise ratio in a CD measurement is, in fact, 0.869 – see Johnson 1996b). Less concentrated solutions may be used if the CD signals are intense and more concentrated samples may, of course, be examined using short path length cuvettes when necessary. The near-UV CD spectra of nucleic acids, which are significantly stronger than those of proteins, should also be recorded with a peak absorption in the range 0.7-1.0.

#### **D. Determination of sample concentration**

Accurate sample concentrations are absolutely essential for the analysis of far-UV CD spectra for secondary structure content and whenever one wishes to make meaningful comparisons between different protein or nucleic acid samples. Lowry or Bradford analyses for proteins are not sufficiently accurate for use with CD measurements unless they have been very carefully calibrated for the protein under investigation using concentrations determined with a more direct method, such as quantitative amino acid analysis. We routinely determine protein and nucleic acid concentrations using absorption spectroscopy. When the extinction coefficient is known the concentration can be calculated with considerable accuracy. The absorption spectrum should ideally be recorded with temperature control and careful attention should be given to correct baseline subtraction, especially when buffers containing reducing agents are being used. Highly scattering samples should always be clarified by low speed centrifugation or filtration prior to concentration determination. If the spectrum still shows significant light scattering, i.e., significant background absorption above  $\sim 315$  nm, a correction must be applied. In most cases it is reasonable to assume that the scattering is Rayleigh in nature and that the absorbance due to scatter is then proportional to  $\lambda^n$  (where the exponent  $n$  is generally close to 4). The light scattering contribution to be subtracted at 280 nm, for example, would then be  $(A_{350\text{nm}})(350/280)^4 = 2.442 \times A_{350\text{nm}}$ . A more elaborate method, easily implemented in a spreadsheet program, is to plot  $\text{Ln}(A_\lambda)$  against  $\text{Ln}(\lambda)$  for  $\lambda > 315$  nm and perform a least-squares fit to the straight line. This method has the advantage that significant deviations from linearity may indicate the presence of contaminants rather than light scattering, and the actual value of the wavelength exponent ( $n$ ) can be calculated. The scattering contribution for the full wavelength range can then be constructed and subtracted from the measured spectrum.

Although it is possible to calculate the extinction coefficient of a protein with reasonable accuracy from its aromatic amino acid content (Gill and von Hippel, 1989; Pace et

al., 1995) it is, of course, much more reliable to measure it. This is best done using the Edelhoch method (see Pace et al., 1995). Make identical dilutions of the protein stock in the experimental buffer and in the same buffer containing 6M guanidine hydrochloride and record absorption spectra with appropriate buffer subtraction. Correct for light scattering if necessary and measure the absorbance at the chosen wavelength. Then, for example, the extinction coefficient at 280 nm is calculated from the amino acid composition as (Pace et al., 1995):

$$\epsilon_{280,buffer} = (A_{280,buffer})(\epsilon_{280,GuHCl}) / (A_{280,GuHCl}) \quad (1)$$

$$\text{where } \epsilon_{280, GuHCl} (\text{M}^{-1}\text{cm}^{-1}) = (\#\text{Trp})(5685) + (\#\text{Tyr})(1285) + (\#\text{cystine})(125). \quad (2)$$

Whatever method of concentration determination is used it is probably wise in any subsequent analyses to assume that it might be in error by up to 5%.

### III. Data Collection

Having chosen the appropriate sample concentration and cuvette path length for the measurement (see section II.C) the user will need to select suitable instrument settings. In addition to choosing the appropriate wavelength range for the measurement it will generally be necessary to select scanning speed, instrumental time constant (or response time), spectral band width, and number of scans to be averaged. Consideration should also be given to the selection of an appropriate temperature for the measurement.

Wavelength range: Far-UV spectra of proteins should generally be scanned from 260 nm to the lowest attainable wavelength. This low wavelength limit will depend largely upon the composition of the buffer being used (see section II.C). Near-UV spectra are routinely scanned over the range 340 to 255 nm for proteins and from 340 nm to the lowest attainable wavelength for nucleic acids.

Scanning speed and time constant (or response time): In the case of analogue instruments, the product of the scanning speed ( $\text{nm}\cdot\text{min}^{-1}$ ) and the time constant (sec) should be less than  $20 \text{ nm}\cdot\text{min}^{-1}\cdot\text{sec}$ . If the instrument uses a response time (equal to three time constants) then the product of scanning speed and response time should be less than  $60 \text{ nm}\cdot\text{min}^{-1}\cdot\text{sec}$  (Johnson 1996b). If significantly higher values are used there will be potentially serious errors in both the positions and intensities of the observed CD bands. Several good general discussions of the potential sources of errors in CD measurements have been published (Hennessey and Johnson 1982; Johnson 1996b; Manning 1989). The signal to noise ratio (S/N) of a CD measurement is proportional to the square root of the number of scans and to the square root of the time constant. Good S/N ratios can therefore be achieved either by averaging multiple fast scans recorded with a short time constant or by recording a small number of slow scans with a long time constant. The choice is largely one of personal preference.

Spectral bandwidth: The spectral bandwidth is generally set to 1 nm but it may occasionally be necessary to use lower values in order to resolve fine structure in near-UV spectra of proteins. Increasing the spectral bandwidth will reduce the noise by increasing light throughput but it should always be 2nm or less in order to avoid distorting the spectrum.

Temperature control: It is good practice to always record CD spectra with temperature control. This is particularly important for the far-UV CD spectra of proteins, which often show quite pronounced temperature dependence, even outside the range of any thermally-induced unfolding of the protein. These small changes in the signal from the folded protein with temperature reflect a true change in conformation and are not simply due to changes in the optical properties of a helix or strand. The changes, which are often linear with temperature, are probably due to fraying of the ends of a helix or to changes in helix-helix interactions (Greenfield, 2004).

Having set the parameters described above it is advisable to run a single scan in order to check that the selection is appropriate. CD spectra will be seriously distorted if the photomultiplier voltage rises above a certain limit, generally of the order of 600 volts. The low wavelength limit selected for far-UV spectra should be reset to a higher value if this photomultiplier voltage limit is exceeded. If the voltage is too high in the near-UV region either the sample concentration or the cuvette path length should be reduced. The near-UV CD spectra of proteins generally have no significant intensity in the 315-340 nm region. If there is a significant signal in this region (often becoming increasingly negative towards lower wavelengths) this may indicate that there is a disulfide contribution to the spectrum (see section IV.B.1). The upper wavelength limit for the scan should then be extended (400 nm is generally sufficient in most cases) in order to allow for correct baseline alignment (see section IV.A).

The full measurement should then be made with enough repeat scans to increase the S/N ratio to acceptable levels. If necessary, for example in performing a titration, make the required additions to the cuvette and repeat the measurement. Making additions (especially to short path length cuvettes) poses several problems. The small volume sample has to be mixed thoroughly, either by inversion or using a long thin pipette tip. This should be done with great care in order to minimise the almost inevitable small loss of solution that will occur during these manipulations. Another frequently encountered problem, especially with dilute protein samples, is that material is lost through absorption onto cuvette walls and pipette tips. In practice there is little to be done about this but it is advisable to check for the problem by performing a 'dummy' titration in which one simply adds buffer to the sample to check for intensity changes that cannot be accounted for solely by dilution. Finally, since additions may increase the total absorption, it is always worth estimating (or better measuring) what the final absorbance will be before beginning an experiment in which additions are to be made.

The final step is to record the baseline using the same cuvette, buffer, and instrument settings. Strictly speaking the baseline should be recorded using a buffer containing something that has the same normal absorption as the sample but no CD signal (Johnson 1996b). However, this is seldom done and is unlikely to be a major problem except with very weak signals. One should not be tempted to reduce the number of scans used to record the baseline, since any noise in the baseline scan will simply be added to the sample scan in subsequent numerical processing.

#### **IV. Data processing and spectral characteristics**

##### **A. Data Processing**

The first step is to subtract the baseline scan from the sample scan. All spectra should have been collected with a starting wavelength (260 nm for far-UV spectra; 340 nm for near-UV spectra -see section III) that gives at least 15-20 nm at the start of the scan where there should be no signal. After baseline subtraction this region should be, and generally is, flat, but the signal may not be zero. The slow vertical drift in signal that is common to CD spectrometers is usually the cause of this problem. The solution is to average the apparent signal over the first 15-20 nm of the spectrum and subtract this average value from the whole of the curve.

The spectrum should be converted to the desired units. In the case of proteins the observed CD signal,  $S$  in millidegrees (Note: 1 millidegree =  $32980 \times \Delta A$ ), is generally converted to either the molar CD extinction coefficient ( $\Delta \epsilon_M$ ) or to the mean residue CD extinction coefficient ( $\Delta \epsilon_{MRW}$ ) using:

$$\Delta \epsilon_M = S / (32980 \cdot C_M \cdot L) \text{ or } \Delta \epsilon_{MRW} = S \cdot MRW / (32980 \cdot C_{mg/ml} \cdot L) \quad (\text{Units : } M^{-1} \text{cm}^{-1}) \quad (3)$$

where  $L$  is the path length (in cm),  $C_M$  is the molar concentration,  $C_{\text{mg/ml}}$  is the concentration in mg/ml and MRW is the mean residue weight (Molecular weight divided by the number of residues). Although large globular proteins generally have a mean residue weight of approximately 111 the actual value must always be calculated in order to avoid potentially large errors in the calculated intensities. Calculating far-UV intensities is almost invariably done on a per residue basis in order to facilitate comparison between proteins and peptides with different molecular weights. Near-UV CD intensities should generally be reported on a molar rather than a per residue basis because only four of the amino acid side chains contribute to the CD signals in this region. In the case of nucleic acids the CD intensities can be calculated using the base, base pair, or molar concentrations.

CD intensities are also sometimes reported as molar ellipticity ( $[\theta]_M$ ) or mean residue ellipticity ( $[\theta]_{\text{mrw}}$ ), which may be directly calculated as:

$$[\theta]_M = S/(10.C_M.L) \text{ or } [\theta]_{\text{mrw}} = S.MRW/(10.C_{\text{mg/ml}}.L) \quad (\text{Units: degrees.cm}^2\text{dmo}^{-1}) \quad (4)$$

$[\theta]$  and  $\Delta\epsilon$  values may be interconverted using the relationship  $[\theta] = 3298.\Delta\epsilon$

## **B. Spectral characteristics**

### **1. Near-UV spectra of proteins**

Near-UV CD bands from individual residues in a protein may be either positive or negative and may vary dramatically in intensity, with residues that are immobilised and/or have coupled-oscillator interactions with neighbouring aromatic residues producing the strongest signals. There is therefore little correlation between the number of aromatic residues and the aromatic CD intensity, and the near-UV CD spectrum of a protein does not generally allow one to say anything explicit about its tertiary structure. Knowledge of the position and intensity of CD bands expected for a particular chromophore is helpful in understanding the

observed near-UV CD spectrum of a protein and the principal characteristics of the four chromophores are therefore summarised below (Strickland 1974; Woody and Dunker 1996):

- Phenylalanine has sharp fine structure in the range 255-270 nm with peaks generally observed close to 262 and 268 nm ( $\Delta\epsilon_M \pm 0.3 \text{ M}^{-1}\text{cm}^{-1}$ ).
- Tyrosine generally has a maximum in the range 275-282 ( $\Delta\epsilon_M \pm 2 \text{ M}^{-1}\text{cm}^{-1}$ ), possibly with a shoulder some 6 nm to the red.
- Tryptophan often shows fine structure above 280 nm in the form of two  $^1L_b$  bands (one at 288 to 293 and one some 7 nm to the blue, with the same sign ( $\Delta\epsilon_M \pm 5 \text{ M}^{-1}\text{cm}^{-1}$ )) and a  $^1L_a$  band (around 265 nm) with little fine structure ( $\Delta\epsilon_M \pm 2.5 \text{ M}^{-1}\text{cm}^{-1}$ ).
- Cystine CD begins at long wavelength ( $> 320 \text{ nm}$ ) and shows one or two broad peaks above 240 nm ( $\Delta\epsilon_M \pm 1 \text{ M}^{-1}\text{cm}^{-1}$ ), the long wavelength peak is frequently negative.

Some of these features are illustrated in Figure 1, which shows near-UV CD spectra of *Drosophila* calmodulin and its complex with a peptide corresponding to the target sequence from CaM-dependent kinase I (Clapperton et al., 2002). This calmodulin contains nine phenylalanines (giving the sharp bands at 262 and 268 nm in panel A), a single tyrosine in the C-terminal domain (giving the broad band around 275 nm in panel A that changes sign upon calcium binding), and no tryptophans. Complex formation with the target peptide introduces a single tryptophan (and 2 further phenylalanines). Panel B shows that the free peptide, which is mobile and unstructured, has only a very weak CD signal but that immobilisation in the complex generates the intense peaks characteristic of tryptophan (see inset to panel B).

## 2. Far-UV spectra of proteins

Far-UV CD spectra of proteins depend upon secondary structure content and simple inspection of a spectrum will generally reveal information about the structural class of the

protein. The characteristic features of the spectra of different protein classes may be summarised as follows (Venyaminov and Yang 1996):

- All- $\alpha$  proteins show an intense negative band with two peaks (at 208 and 222 nm) and a strong positive band (at 191 - 193 nm). The intensities of these bands reflect alpha-helical content.  $\Delta\epsilon_{\text{mrw}}$  values for a totally helical protein would be of the order of  $-11 \text{ M}^{-1}\text{cm}^{-1}$  (at 208 and 222 nm) and  $+21 \text{ M}^{-1}\text{cm}^{-1}$  (at 191 - 193 nm).
- The spectra of regular all- $\beta$  proteins are significantly weaker than those of all- $\alpha$  proteins. These spectra usually have a negative band (at 210-225 nm,  $\Delta\epsilon_{\text{mrw}}$   $-1$  to  $-3.5 \text{ M}^{-1}\text{cm}^{-1}$ ) and a stronger positive band (at 190-200 nm,  $\Delta\epsilon_{\text{mrw}}$   $2$  to  $6 \text{ M}^{-1}\text{cm}^{-1}$ ).
- Unordered peptides and denatured proteins have a strong negative band (at 195 - 200 nm,  $\Delta\epsilon_{\text{mrw}}$   $-4$  to  $-8 \text{ M}^{-1}\text{cm}^{-1}$ ) and a much weaker band (which can be either positive or negative) between 215 and 230 nm ( $\Delta\epsilon_{\text{mrw}}$   $+0.5$  to  $-2.5 \text{ M}^{-1}\text{cm}^{-1}$ ).
- $\alpha+\beta$  and  $\alpha/\beta$  proteins almost always have spectra dominated by the  $\alpha$ -helical component and therefore often show bands at 222, 208 and 190 - 195 nm. In some cases there may be a single broad minimum between 210 and 220 nm because of overlapping  $\alpha$ -helical and  $\beta$ -sheet contributions.

Some of these features are illustrated in Figure 2, which shows the far-UV CD spectra of myoglobin (4mbn.pdb), prealbumin (2pab.pdb), and acid denatured staphylococcal nuclease at pH 6.2 (6°C). These spectra and those of many other proteins are available on the internet (at <http://lamar.colostate.edu/~sreeram/CDPro/>).

### 3. Nucleic acids

Because the aromatic bases themselves are planar they do not possess any intrinsic CD signals; it is the presence of the sugars that creates the asymmetry that leads to the small CD signals of the monomeric nucleotides. Likewise, the stacking of the bases in the different

polymeric forms results in the close contact and electronic interactions that produce the intense CD signals of the nucleic acids and oligonucleotides. CD is sensitive to secondary structure because the precise nature of these interactions determines the shape of the spectrum. The principal conformational forms of the nucleic acids are the A- and B-forms. In neutral aqueous buffers at moderate salt concentrations DNA is usually in the B-form, whilst RNA adopts the A-form. These conformations have characteristic CD spectra that depend on base composition and somewhat on sugar type (Johnson 1996a). Variation in spectral shape with base composition is, of course, significantly more important with short oligonucleotides. For this reason CD is usually used empirically in the study of nucleic acid conformation by comparing an experimentally measured spectrum to that of a nucleic acid of known structure (Johnson 1996a).

## V Applications

### A. Secondary structure content of proteins

The estimation of protein secondary structure content from far-UV CD spectra is one of the most widely used applications of CD. If the chiroptical properties of the different secondary structure elements are assumed to be additive then the measured far-UV CD spectrum of a protein at any wavelength ( $\lambda$ ) can be expressed as:

$$S(\lambda) = \sum_{i=1}^n f_i S_i(\lambda) \quad (5)$$

where  $S(\lambda)$  is the observed mean residue CD,  $f_i$  is the fraction of the  $i^{\text{th}}$  conformation, and  $S_i(\lambda)$  is the reference CD signal for the  $i^{\text{th}}$  conformation. If the reference spectra are known then the  $f_i$ 's can be solved from a series of simultaneous equations (one for each  $\lambda$ ) by a least squares method. The original approach to this problem used synthetic polypeptides as model

compounds to provide reference spectra for the  $\alpha$ -helix,  $\beta$ -form, and unordered form (Greenfield and Fasman, 1969). The approach was later extended by Brahms and Brahms (1980) to include a reference for  $\beta$ -turns. An alternative approach is to compute the appropriate reference spectra from the spectra of proteins of known secondary structure using Eq. (5). Saxena and Wetlaufer (1971) used spectra of just three proteins to derive reference spectra for  $\alpha$ -helix,  $\beta$ -form, and unordered form. The same reference spectra were subsequently derived from five proteins (Chen et al., 1972) and then eight proteins (Chen et al., 1974). Chang et al., (1978) used 15 proteins and added a general term for  $\beta$ -turn. In addition to the assumption of additivity (see Eq. (5)) these methods depend on four other major assumptions (Sreerama et al., 2000): (1) that the effect of tertiary structure on CD is negligible, (2) that the ensemble-averaged solution structure is equivalent to the X-ray structure, (3) that the contributions from non-peptide chromophores do not influence the analysis, and (4) that the geometric variability of secondary structures need not be explicitly considered. Although assumptions (1) and (2) are probably valid, assumptions (3) and (4) are probably invalid in many cases (Sreerama et al., 2000). For example, several recent experiments, particularly those using site-directed mutagenesis, have shown that aromatic residues can make surprisingly large contributions to the far-UV CD spectra of some proteins and this obviously has serious implications for secondary structure estimation (see Woody and Dunker (1996) and references therein). Assumption (4) is even more likely to be incorrect because it is highly unlikely that the secondary structure of a protein will be ideal. Thus, for example, a single reference spectrum for the  $\alpha$ -helix cannot account for variable contributions from regular helices,  $3_{10}$ -helices and distorted helices, and the question of chain-length dependence of the  $\alpha$ -helix intensity (see Chen et al., 1974) is generally not explicitly considered. Likewise, a single reference spectrum for the  $\beta$ -form cannot account for variable contributions from parallel and anti-parallel forms, sheet width and chain length dependence

of the CD signal, and the fact that this form is often twisted and non-planar, with a degree of twisting that varies from protein to protein.

Two approaches designed to circumvent the problem of defining suitable reference spectra were introduced in 1981. Provencher and Glockner (1981) developed a program (CONTIN), which uses ridge regression analysis to fit the spectrum of the unknown directly by a linear combination of the spectra of proteins with known structure. In this program, the contribution of each reference spectrum is kept small unless it contributes to a good fit to the raw data. In the second approach, introduced by Hennessey and Johnson (1981), singular value decomposition (SVD) was applied to extract orthogonal basis curves from a set of spectra of proteins with known structure. The deconvoluted basis curves, each of which has a unique shape, can be associated with a known mixture of secondary structures and can therefore be used to analyze the conformation of an unknown sample. Convex constraint analysis (CCA) has also been used to extract basis spectra (Perczel et al., 1991).

The reference or basis curves obtained from linear regression, SVD, or CCA will depend strongly on the proteins included in the data set and an inadequate representation of the spectrum of the unknown will result in a failed analysis (Sreerama and Woody, 2004a). For this reason, most of the methods currently employed achieve greater flexibility by using some form of selection process to create several different subsets of reference spectra in an attempt to include proteins that are important for the analysis and exclude those that have an adverse effect. This approach, known as variable selection, is implemented in different ways by different programs.

The basic method (implemented in VARSLC; Manavalan and Johnson 1987) combines variable selection with SVD in the following way. Starting from a large database of reference proteins the program sequentially eliminates one or more spectra to create a large series of new databases. SVD is used on these reduced data sets to evaluate the unknown

conformation; all the results are then examined and those fulfilling certain selection criteria (see Sreerama and Woody, 2004a) for a good fit are averaged. The two other approaches for introducing variable selection with SVD are known as the minimal basis and locally linearized approaches. In the minimal basis approach (CDSSTR; Johnson 1999) small subsets of reference spectra are randomly selected to create the minimal basis. This creates a large number of different combinations and the process is stopped when an acceptable number of valid solutions has been obtained. In the locally linearized approach spectra are systematically removed from the reference set on the basis of their similarity to the analyzed spectrum, with the least similar reference spectra being removed first. Local linearization is employed in SELCON (Sreerama and Woody 1993), which uses a modification of the variable selection method called the self-consistent method. The program includes the spectrum of the unknown in the database and makes an initial guess of the structure. The database is decomposed using SVD and the unknown is analyzed. The solution replaces the initial guess and the process is repeated until the results are self-consistent. Variable selection has also been explicitly introduced into CONTIN by Sreerama and Woody (2000b) in the locally linearized implementation known as CONTIN/LL. Finally, neural networks are programs that are used to detect patterns in data, and two such programs have been described for the analysis of far-UV CD data of proteins (K2D; Andrade et al., 1993 and CDNN; Bohm et al., 1992) but these do not include variable selection

The various methods available have been extensively discussed in the literature (Johnson 1990; Greenfield 1996, 2004; Sreerama and Woody 1993, 1994, 2000a,b, 2004a; Van Stokkum et al. 1990; Venyaminov and Yang 1996) and Greenfield (2004) has made useful recommendations for selecting the program that is most appropriate for a particular purpose. Two important points are worth remembering (Manavalan and Johnson, 1987; Venyaminov and Yang, 1996): these are that CD analysis methods cannot recognize a failed analysis and

that the solution selected by a particular program may not necessarily be the best of those examined.

Three very popular programs (SELCON, CDSSTR, and CONTIN/LL) provided with several reference sets with different wavelength ranges are available on the internet (at <http://lamar.colostate.edu/~sreeram/CDPro/>) with the CDPro software package (a more extensive list of websites where CD analysis programs are available is given in Greenfield, 2004). The question of whether it is appropriate to use the standard soluble protein reference data sets for analyzing the CD spectra of membrane proteins has been addressed by Wallace et al. (2003) and Sreerama and Woody (2004b). The latter authors have shown that the data sets available with the CDPro package do perform reasonably well and that this performance is improved still further by including membrane proteins in the reference set. Finally, methods available for obtaining additional structural information by calculating the number of secondary structure segments in a protein and identifying its tertiary structure class have been described (see Sreerama and Woody (2004a) and references therein).

## **B Detecting altered conformation**

### **1. Solution conditions**

The stability of any particular secondary structure element in a protein or nucleic acid will depend upon several different factors. For example, the stability of an  $\alpha$ -helix is determined by the number and strength of interhelical hydrogen bonds, by interactions (potentially either stabilizing or destabilizing) between ionized side chains, and by interactions with other secondary structural elements in the protein. These factors, and therefore the protein's structure are often influenced by changes in solution conditions. For example, changes in the pH and ionic strength of the solution will generally affect interactions between charged side chains in proteins and this can lead to significant changes in secondary

and/or tertiary structure. Such changes are readily studied using CD. For example, CD has been extensively used in the study of conformational transitions in peptides (see Figure 3) and in studies of the  $\alpha$ -helix to  $\beta$ -sheet conversion in recombinant prion proteins induced by the addition of low concentrations of denaturants and/or salt at low pH (Morillas et al., 2001; Swietnicki et al., 2000).

Alcohols with low dielectric constants, such as trifluoroethanol (TFE) and hexafluoroisopropanol, are often used to promote the formation of stable conformations in peptides that are unstructured in aqueous solution. There may also be a significant effect of even relatively simple alcohols on the structure of a protein (Griffin et al., 1986). The low dielectric constant of these structure-inducing co-solvents is thought to resemble that of the interior of a protein. Although this low dielectric environment presumably strengthens favorable electrostatic interactions this particular effect may be partially offset by increased counterion binding. The major effect is generally attributed to weaker hydrogen bonding of the amide protons to the solvent with concomitant strengthening of intramolecular hydrogen bonds and stabilization of the helix.

The effects high temperature will be discussed further in section V.C.2. CD spectra can, when necessary, be recorded at temperatures below zero by using suitable water-alcohol or water-glycerol mixtures. It is essential in such studies to check for any direct effect of the solvent itself on the conformation of the protein. This is done by measuring the CD spectrum in the solvent at room temperature and comparing it with the spectrum measured in an aqueous buffer at this temperature.

## **2. Changes accompanying complex formation**

Protein-protein and protein-nucleic acid interactions are often accompanied by changes in the intrinsic CD of one or both of the components owing to changes in secondary

structure (Greenfield and Fowler 2002) and/or the environment of aromatic groups (Clapperton et al., 2002). Such changes may also be caused by the binding of small ligands, such as drugs and metal ions (Martin and Bayley, 1986), to macromolecules, and in certain cases the ligand itself may change its optical activity or become optically active. Because CD is a quantitative technique these changes can be used to determine equilibrium dissociation constants for complex formation (see section V.D). The source of the change in CD is immaterial for this application. Even when it is not possible to use these changes to determine dissociation constants they do provide unequivocal evidence that two species interact, and frequently provide useful information about the nature of the interaction. For example, stoichiometric titrations of calmodulin with metal ions, in which the signal from tyrosine-138 in the C-terminal domain (see Figure 1A) is monitored, can be used to establish the preference of different metals for binding to the individual calmodulin domains (see Figure 4).

### **3. Structural analysis of recombinant native proteins and their mutants**

When a wild-type protein is produced using recombinant DNA technology it is important to demonstrate that it has the same overall structure as a protein isolated by traditional methods. This is especially important if inclusion body formation necessitated the addition of denaturation and refolding steps to the isolation procedure. Comparison of the near- and far-UV CD spectra is ideal for this purpose.

When working with mutant proteins it is, of course, good practice to test for any effect of the mutation on the general conformation the protein and, here again, CD provides a convenient means of doing this with the limited amounts of material that are sometimes available. A significant difference in shape between the far-UV CD spectra of the wild-type and mutant proteins can be an indication that the mutation has produced some change in the secondary structure. However, this may not be true if the mutation involves the introduction

or replacement of an aromatic residue (Tyr or Trp) because these residues may, in certain cases, contribute significantly to the far-UV spectrum (see section V.A). A change in shape may also occur if the mutation destabilizes the protein to such an extent that it exists as a mixture of folded (with wild-type conformation) and unfolded forms under the conditions of the measurement. This is a distinct possibility with small proteins, which frequently have relatively low free energies for unfolding and are therefore only just stable at room temperature (see section V.C.2). When a difference in shape is observed, especially if the mutant looks to be less well folded, it is useful to repeat the measurements at a lower temperature, where the protein would be expected to be more stable (see section V.C.2). If the difference between the wild-type and mutant spectra is smaller at the lower temperature then it is highly likely that the mutation has had a significant destabilizing effect and the stability properties should be fully characterized using one of the methods described in section V.C.

Changes in the near-UV region of the spectrum may indicate that the mutation has caused a change in tertiary structure but could again result from destabilization of the protein. In addition it is important to remember that significant changes may often derive from rather subtle changes in the environment of one or more of the aromatic residues that are not necessarily associated with any major structural change. Mutations that involve the introduction or removal of tyrosine or tryptophan would, of course, be expected to change the spectrum and serial replacement of aromatic residues by site directed mutagenesis can be used to investigate the contributions of individual aromatic residues to the near-UV CD spectra of proteins (Craig et al., 1989).

It is generally wise to be suspicious if the wild-type and mutant spectra differ in intensity but not in shape. If the spectra can be superimposed by applying a simple multiplication factor then there has most probably been an error in concentration

determination (see section II.D), particularly if the same factor serves to superimpose both the near- and far-UV spectra.

### **C. CD in the study of protein stability**

Unfolding of macromolecules is generally studied by using an optical method (absorbance, fluorescence, or CD) or by using differential scanning calorimetry (DSC). CD is very widely employed in the study of protein stability because unfolding is almost invariably accompanied by major changes in both the near- and far-UV CD spectra. The unfolding mechanism for many small globular proteins is essentially a two-state process, where only the folded (F) and unfolded (U) states are significantly populated at equilibrium. Conformational stability is defined in terms of the free energy of unfolding:

$$\Delta G = -RT \ln K \quad (6)$$

where K is the unfolding constant ( $= [U]/[F]$ ), R is the universal gas constant, and T is the absolute temperature.

The free energy of unfolding is determined by monitoring an appropriate CD signal as a function of the concentration of a chemical denaturant (urea or guanidine hydrochloride) or as a function of temperature. Because the object is to measure thermodynamic parameters the unfolding must be reversible and this must be confirmed by demonstrating that the signal of the folded protein is recovered upon lowering the temperature or upon removing the denaturant. It is also essential that equilibrium is reached at each point in the unfolding curve. In general, the nearer the midpoint of the transition the longer will be the time taken to reach equilibrium.

#### **1. Chemical Denaturation**

Chemical denaturation experiments are generally performed in one of two ways: (1) By measuring the CD of a large number of individually prepared samples, each with a different concentration of denaturant but the same concentration of protein; or (2) By titrating a solution of protein with concentrated denaturant containing protein at the same concentration. This approach can, of course, only be used when the unfolding is reversible. The titration can also be performed in the reverse direction; i.e. adding protein in buffer to protein in concentrated denaturant. In most cases one only requires values of CD intensity at a single wavelength for the analysis. To allow proper baseline alignment at higher wavelength, these values should be taken from full wavelength scans whenever the signal is weak. When the signal is strong, the baseline alignment problems should be minimal and observation at a single wavelength may be adequate. For far-UV titrations it may be helpful to use a solution at the normal concentration for a far-UV measurement (i.e., 0.1-0.15 mg/ml) but use a longer path length cuvette. This restricts the accessible lower wavelength range, but normally permits measurements in the region of interest (generally 220nm).

With both of the approaches described above one obtains the variation of the CD signal with denaturant concentration. It is quite common to observe that the CD signal of the folded and unfolded forms show some dependence on denaturant concentration,  $[D]$ , outside the region of the unfolding transition. The variation of the CD signal,  $S$ , with  $[D]$  must therefore be described by:

$$S = (S_F + \beta_F [D]) \cdot F_F + (S_U + \beta_U [D]) \cdot F_U \quad (7)$$

where  $S_F$  and  $\beta_F$  are the intercept and slope of the pre-transition baseline,  $S_U$  and  $\beta_U$  are the intercept and slope of the post-transition baseline, and  $F_F$  and  $F_U$  are the fractions of protein present in the folded and unfolded forms. The dependence of the free energy of unfolding on  $[D]$  is most often described using the linear extrapolation model,

$$\Delta G = \Delta G(H_2O) + m[D] \quad (8)$$

where  $\Delta G(\text{H}_2\text{O})$  is the free energy of unfolding in the absence of denaturant and  $m$  describes the dependence on the denaturant concentration. Other extrapolation models may be used but they will not be described here. For simple two-state unfolding  $K = [U]/[F] = F_U/F_F$ .

Substituting  $F_F = 1/(1+K)$  and  $F_U = K/(1+K)$  along with  $K = e^{-\Delta G/RT}$  gives:

$$S = \frac{(S_F + \beta_F[D]) + (S_U + \beta_U[D])e^{\frac{-\Delta G(\text{H}_2\text{O}) - m[D]}{RT}}}{1 + e^{\frac{-\Delta G(\text{H}_2\text{O}) - m[D]}{RT}}} \quad (9)$$

and analysis using standard non-linear least-squares methods should yield  $\Delta G(\text{H}_2\text{O})$  and the  $m$  value. Guanidine hydrochloride is a more powerful denaturant than urea because it has a larger  $m$  value for any particular protein (Myers et al. (1995) have provided a useful compilation of  $m$  values for a large number of proteins).

## 2. Thermal unfolding

In thermal unfolding experiments the CD signal at a single wavelength is recorded whilst the temperature is slowly increased (preferably at no more than 1°/min). This is generally done using a Peltier unit or a programmable water bath. In either case it is advisable to check the actual temperature profile of the device by measuring the temperature in the cuvette using an immersible electronic probe. Buffers with high thermal coefficients (e.g., Tris) should be avoided if possible in order to avoid large changes in pH as the temperature is increased.

The CD signal of the folded and unfolded forms will generally show some dependence on temperature so that there will be pre- and post-transition slopes, as with chemical denaturation. The variation of the CD signal,  $S$ , with temperature must therefore be described by:

$$S = (S_F + \beta_F T) \cdot F_F + (S_U + \beta_U T) \cdot F_U \quad (10)$$

where the  $S$  and  $\beta$  values describe the slopes and intercepts of the pre- and post-transition slopes. The variation of  $\Delta G$  with temperature is described by the modified Gibbs-Helmholtz equation (Becktel and Schellman, 1987):

$$\Delta G = \Delta H_{T_m} \left( 1 - \frac{T}{T_m} \right) + \Delta C_p \left\{ (T - T_m) - T \ln \left( \frac{T}{T_m} \right) \right\} \quad (11)$$

where  $T_m$  and  $\Delta H_{T_m}$  are the mid-point melting temperature and the enthalpy at  $T_m$  respectively, and  $\Delta C_p$  is the difference in heat capacity between the folded and unfolded protein (see Robertson and Murphy, 1997).

Substituting  $F_F = 1/(1+K)$  and  $F_U = K/(1+K)$  along with  $K = e^{-\Delta G/RT}$  gives:

$$S = \frac{(S_F + \beta_F T) + (S_U + \beta_U T) e^{-\Delta H_{T_m} \left( 1 - \frac{T}{T_m} \right) - \Delta C_p \left\{ (T - T_m) - T \ln \left( \frac{T}{T_m} \right) \right\}}}{1 + e^{-\Delta H_{T_m} \left( 1 - \frac{T}{T_m} \right) - \Delta C_p \left\{ (T - T_m) - T \ln \left( \frac{T}{T_m} \right) \right\}}} \quad (12)$$

Although it is theoretically possible to fit the CD data to this equation there is generally not enough information to determine  $\Delta C_p$  accurately (because it is the second derivative of  $\Delta G$  with respect to temperature). Greenfield (2004) has noted that trying to include the  $\Delta C_p$  term can, in fact, actually worsen the agreement between  $T_m$  values determined by CD and DSC. It is therefore common practice to set  $\Delta C_p$  equal to zero in the analysis. If  $\Delta C_p$  can be determined using another method, such as DSC, then  $\Delta H_{T_m}$  and  $T_m$  can be used to calculate  $\Delta G$  at any temperature using the modified Gibbs-Helmholtz equation (Eq. (11)). The variation of the free energy for unfolding of proteins with temperature has several interesting features (see Fig. 5).  $\Delta G^0$  will not vary linearly with temperature because the  $\Delta C_p$  term is large and positive for protein unfolding (Robertson and Murphy 1997). In addition to the well-characterized high temperature unfolding (Midpoint =  $T_m$ , where  $\Delta G^0 = 0$ ) there will also be a temperature of maximum stability, below which the protein will again

become less stable (curves **b** and **c**). In certain cases the free energy for unfolding will become small enough at low temperatures for one to observe the phenomenon of cold-induced denaturation (curve **c**), which is easily monitored using CD (Chen and Schellman, 1989). Another feature illustrated in Fig. 5 is that proteins with a high  $T_m$  for high temperature unfolding do not necessarily have high overall stability at room temperature (compare curves **a** and **b**). Hence, the relative stability of two different proteins at ambient temperatures cannot be estimated solely from  $T_m$  measurements; a full analysis yielding  $\Delta H_{T_m}$  and  $\Delta C_p$  values is required for this (using Eq. (11)).

### **3. Unfolding intermediates.**

The two-state nature of the unfolding process can often be confirmed by monitoring the unfolding using different probes. If the unfolding is truly two-state then the transition should be identical whether monitored by near-UV CD (reporting on tertiary structure) or far-UV CD (reporting on secondary structure). Under some conditions (generally mildly denaturing conditions such as low pH) some proteins, for example  $\alpha$ -lactalbumin (Kuwajima, 1996) form an intermediate molten globule state that is neither fully folded nor fully unfolded (see Arai and Kuwajima (2000) for a review of the role of the molten globule state in protein folding). This intermediate state retains most of the native secondary structure and generally has a native-like far-UV CD spectrum, but has a highly fluid tertiary structure and therefore has a near-UV CD spectrum characteristic of the fully unfolded form.

Some proteins, particularly those containing more than a single domain, may not unfold in a single transition. In such cases, the data can be fit to a model in which it is assumed that the unfolding is represented by the sum of two or more independent transitions. The equation used to fit the data is then the appropriate sum of those for the individual transitions given above (see, for example, Masino et al., 2000).

## D. Determination of equilibrium dissociation constants

### 1. Direct Methods

As noted in section V.B.2 the interaction of a macromolecule with another macromolecule or small molecule is often associated with a change in the CD signal of one or both of the components. Although not generally the method of choice owing to its relatively poor S/N ratio, these changes can sometimes be used to determine equilibrium dissociation constants. Whilst a detailed discussion of the methods available for characterizing such interactions is beyond the scope of this chapter (see Greenfield (1975, 2004) for general reviews), it is appropriate to make some remarks here. We will do this with reference to the simplest possible binding model, formation of a 1:1 complex according to the following scheme.



The primary requirement for a direct titration, as with any spectroscopic method, is that the sum of the spectra of the components (A and B) should be different from that of the complex (AB) at some wavelength (i.e.,  $\Delta\epsilon_A + \Delta\epsilon_B \neq \Delta\epsilon_{AB}$ ). If this requirement is fulfilled then a titration of A with B (or B with A; see below) can, in suitable cases, allow determination of the dissociation constant,  $K_d$ . The observed CD signal,  $S$ , for any mixture of A and B is given by:

$$\text{Signal} = [A]\Delta\epsilon_A + [B]\Delta\epsilon_B + [AB]\Delta\epsilon_{AB} \quad (14)$$

Eliminating  $[A]$  and  $[B]$  using the mass balance relationships ( $[A] = A_T - [AB]$  and  $[B] = B_T - [AB]$ ) and substituting the expression the following expression for  $[AB]$

$$[AB] = \frac{([A_T] + [B_T] + K_d) - \left( ([A_T] + [L_T] + K_d)^2 - 4[A_T][B_T] \right)^{0.5}}{2} \quad (15)$$

gives

$$\text{Signal} = \Delta\epsilon_A A_T + \Delta\epsilon_B B_T + (\Delta\epsilon_{AB} - \Delta\epsilon_A - \Delta\epsilon_B) \left( \frac{(A_T + B_T + K_d) - \left( (A_T + B_T + K_d)^2 - 4A_T B_T \right)^{0.5}}{2} \right) \quad (16)$$

where  $A_T$  and  $B_T$  are the total concentrations of A and B. Because  $\Delta\epsilon(A)$  and  $\Delta\epsilon(B)$  can be measured separately there are only two unknowns in this equation, the  $K_d$  and  $\Delta\epsilon(AB)$ .

Several important factors must be considered in designing the experiment.

(1) Titration mode: The titration can, of course, be performed by adding A to B or by adding B to A. The choice may be dictated by availability and/or solubility of one of the components.

If both components have a CD signal then the component being added should have the weaker signal of the two as this permits more accurate determination of the end point. For example, in studies of protein-nucleic acid interactions the signal from the protein will generally be much weaker than that of the nucleic acid and the titration is best performed by adding the protein to the nucleic acid (Gray 1996). The other factor to be considered here is that the total absorbance at the observation wavelength (see below) must be kept below 1.2.

(2) Choice of wavelength. In the simplest cases the wavelength selected is where the signal change is largest. However, two factors may dictate otherwise. First, particularly in far UV measurements, it may be advisable to sacrifice amplitude for improved S/N ratio at another wavelength. Second, in experiments with strongly absorbing samples moving away from the absorption maximum will permit higher concentrations to be used.

(3) Concentration range. It is clear from inspection of Eq. (16) that a successful analysis requires that the variation of the CD signal with concentration accurately defines both  $K_d$  and  $\Delta\epsilon(AB)$ . In addition it must be possible to demonstrate that it is indeed a 1:1 complex that is being formed. Consider for example, a titration of A with B performed at a constant concentration of A ( $= A_T$ ) by ‘spiking’ the stock solution of B with A (at  $A_T$ ). One needs to select an appropriate value for  $A_T$  and estimate what final concentration of B ( $= B_{T(\text{Final})}$ ) will be required. The following discussion is based on the assumption that a preliminary

examination has allowed an approximate value of the dissociation constant to be determined. The estimate of  $B_{T(\text{Final})}$  is easy to deal with since  $\Delta\varepsilon(\text{AB})$  can clearly only be determined accurately if a large fraction of A (ideally  $> 0.8$ ) is saturated with B at the end of the titration. Fractional saturation of A ( $= f_A$ ) is related to the concentration of unbound B by the relationship  $f_A = [B]/(K_d + [B])$ ; therefore  $[B]$  at the end of the titration should be  $> 5 \times K_d$ . The minimum total concentration of B required is then calculated as  $B_{T(\text{Final})} = (0.8 \times A_T + 5 \times K_d)$ . The experiment should be designed with this minimum concentration in mind but in practice it is, of course best to reach the highest concentration possible. This is especially true if one is not using 'spiking' as dilution during the titration can lead to the appearance of a false plateau and the titration can be ended prematurely.

The choice of an appropriate value of  $A_T$  can be illustrated by inspection of Fig. 6. This figure shows  $f_A$  as a function of the  $B_T/A_T$  ratio for a system with  $K_d = 1 \mu\text{M}$  and different values of  $A_T$ . In the ideal case (Curve a,  $A_T = 2 \times K_d$ ) the  $K_d$  and the  $\Delta\varepsilon(\text{AB})$  values will be well defined and the shape of the binding curve will clearly be consistent with 1:1 stoichiometry. If  $A_T$  is too high one observes what is essentially a stoichiometric titration with a very sharp transition at the 1:1 point (Curve b,  $A_T = 25 \times K_d$ ) and the  $K_d$  cannot be accurately determined because there will be no point during the titration where A, B, and AB coexist at significant levels. In practical terms this means that CD measurements require that  $A_T < 5-10 \times K_d$ , depending on the signal to noise ratio. It should be noted however that titrations performed at high concentrations are useful for demonstrating 1:1 stoichiometry. If  $A_T$  is too low then the binding curve will be very flat (Curve c,  $A_T = 0.2 \times K_d$ ) and even if  $\Delta\varepsilon(\text{AB})$  can be determined it will not be possible to verify that the stoichiometry is 1:1. In practical terms this means that CD measurements should ideally be performed with  $A_T > 0.5 \times K_d$ . Although this range of  $A_T$  values is relatively narrow the availability of cuvettes with

different path lengths means that a large range of  $K_d$  values can, at least in principle, be determined using CD methods.

## 2. Indirect methods

This section contains a brief survey of the methods that can be used when there is no suitable signal change accompanying the interaction.

When a ligand binds to a protein only (or preferentially) when it is in the folded form then the protein will be more stable in the presence of the ligand. This is because the folded and unfolded forms are in equilibrium; complex formation removes some of the folded form and some of the unfolded protein must refold to maintain the equilibrium. The increase in stability in the presence of the ligand can be used to determine the equilibrium dissociation constant. In the case of thermal unfolding experiments, the effect of ligand on the mid-point for unfolding can be used. If the effect of ligand binding on the enthalpy for unfolding is small, then (Schellman 1975):

$$\Delta T_m = \frac{T_m T_m^* R}{\Delta H_{T_m^*}} \ln \left( 1 + \frac{[L]}{K_d} \right) \quad (17)$$

where  $\Delta T_m$  is the change in the midpoint of the unfolding curve,  $T_m^*$  and  $T_m$  are the midpoint values in the presence and absence of ligand,  $\Delta H_{T_m^*}$  is the enthalpy of unfolding in the presence of the ligand at  $T_m^*$ , and  $[L]$  is the free ligand concentration. A disadvantage of this approach is that one gets an estimate of the  $K_d$  at an elevated temperature. A more detailed evaluation of this approach has been described recently by Mayhoo and Windsor (2005).

The change in stability of a protein to chemical denaturation can be used in the same way by measuring the free energies of unfolding in the presence and absence of the ligand (Pace and McGrath 1980). The relevant equation here is:

$$\Delta G_L - \Delta G = RT \cdot \text{Ln} \left( 1 + \frac{[L]}{K_d} \right) \quad (18)$$

where  $\Delta G_L$  and  $\Delta G$  are the free energies of unfolding in the presence and absence of the ligand, respectively. When the interaction involves charged species these experiments are better performed with urea rather than guanidine hydrochloride because of the high ionic strength of the latter. If the interaction does not have simple 1:1 stoichiometry then the expressions are more complicated (see, for example, Masino et al., 2000). It should be emphasized that this type of experiment is useful even if precise constants cannot be determined because a change in stability is an unequivocal demonstration that an interaction does exist.

Another method that can sometimes be used to determine a dissociation constant for a ligand that binds with no observable signal change is to use a displacement (or competition) assay in which the optically ‘silent’ ligand is used to displace a ligand that does give a signal. One major advantage of this approach is that it depends on the difference in affinity between the two ligands, and not on the absolute values of the affinities. It is therefore possible to use this approach to study very high affinity interactions (Martin and Bayley, 2002).

### **Acknowledgements**

S.R.M. is grateful to Dr Peter Bayley (N.I.M.R.) for introducing him to circular dichroism and for his encouragement and support. The preparation of this review was supported in part by Wellcome Trust Grant 072930/Z/03/Z to M.J.S.

## References

- Andrade, M.A., Chacón, P., Merelo, J.J. and Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* 6, 383-390.
- Arai, M., and Kuwajima, K. (2000). Role of the molten globule state in protein folding. *Adv. Protein Chem.* 53, 209-282.
- Becktel, W.J., and Schellmann, J.A. (1987). Protein stability curves. *Biopolymers* 26, 1859-1877.
- Böhm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* 5, 191-195.
- Brahms, S., and Brahms, J. (1980). Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* 138, 149-178.
- Chang, C.T., Wu, C.-S.C., and Yang, J.T. (1978). Circular dichroic analysis of protein conformation: Inclusion of the  $\beta$ -turns. *Anal. Biochem.* 91, 13-31
- Chen, B.-L., and Schellman, J.A. (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. *Biochemistry* 28, 685-691.
- Chen, Y.-H., Yang, J.T., and Martinez, H.M. (1972). Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* 11, 4120-4131.
- Chen, Y.-H., Yang, J.T., and Chau, K.H. (1974). Determination of the helix and  $\beta$  form of proteins in aqueous solution by circular dichroism. *Biochemistry* 13, 3350-3359.
- Craig, S., Pain, R.H., Schmeissner, U., Virden, R., and Wingfield, P.T. (1989). Determination of the contributions of individual aromatic residues to the CD spectrum of IL-1 beta using site directed mutagenesis. *Int. J. Pept. Protein Res.* 33, 256-262.

Clapperton, J.A., Martin, S.R., Smerdon, S.J., Gamblin, S.J., and Bayley, P.M. (2002). Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. *Biochemistry* 41, 14669-14679.

Fasman, G.D. (1996). Differentiation between transmembrane helices and peripheral helices by deconvolution of circular dichroism spectra of membrane proteins. In "Circular dichroism and the conformational analysis of biomolecules" (G.D. Fasman, ed.), pp. 381-412. Plenum Press, New York.

Gill, S.C. and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319-326.

Gray, D.M. (1996). Circular dichroism of protein-nucleic acid interactions. In "Circular dichroism and the conformational analysis of biomolecules" (G.D. Fasman, ed.), pp. 469-500. Plenum Press, New York.

Greenfield, N.J. (1975). Enzyme ligand complexes: spectroscopic studies. *CRC Crit. Rev. Biochem.* 3, 71-110.

Greenfield, N.J. (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* 235, 1-10.

Greenfield, N. J. (2004). Analysis of circular dichroism data. *Meth. Enzymol.* 383, 282-317.

Greenfield, N.J., and Fasman, G.D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* 8, 4108-4116.

Greenfield, N.J., and Fowler, V.M. (2002). Tropomyosin requires an intact N-terminal coiled coil to interact with tropomodulin. *Biophys. J.* 82, 2580-2591.

Griffin, M.C.A., Price, J.C., and Martin, S.R. (1986). The effect of alcohols on the structure of caseins. Circular dichroism studies of kappa- casein A. *Int. J. Biol. Macromol.* 8, 367-371.

Johnson, W. C., jr (1985). Circular dichroism and its empirical application to biopolymers. *Methods Biochem. Anal.* 31, 61-163.

Hennessey, J. P., jr., and Johnson, W. C., jr. (1981). Information content in the circular dichroism spectra of proteins. *Biochemistry* 20, 1085-1094.

Hennessey, J. P., jr., and Johnson, W. C., jr. (1982). Experimental errors and their effect on analyzing circular dichroism spectra of proteins. *Anal. Biochem.* 125, 177-188.

Johnson, W. C., jr. (1988). Secondary structure of proteins through circular dichroism spectroscopy. *Ann. Rev. Biophys. Biochem.* 17, 145-166.

Johnson, W. C., jr. (1990). Protein secondary structure and circular dichroism: a practical guide. *Proteins: Struct. Funct. Genet.* 7, 205-214.

Johnson, W.C. Jr. (1996a). Determination of the conformation of nucleic acids by electronic CD. In "Circular dichroism and the conformational analysis of biomolecules" (G.D. Fasman, ed.), pp. 433-468. Plenum Press, New York.

Johnson, W. C., jr. (1996b). Circular dichroism instrumentation. In "Circular dichroism and the conformational analysis of biomolecules" (G.D. Fasman, ed.), pp. 635-652. Plenum Press, New York.

Johnson, W.C., Jr. (1999). Analysing protein circular dichroism spectra for accurate secondary structure. *Proteins: Struct. Funct. Genet.* 35, 307-312.

Kuwajima K (1996). The molten globule state of alpha-lactalbumin. *FASEB J.* 10, 102-9.

Manavalan, P., and Johnson, W.C., jr. (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* 167, 76-85.

Manning, M. C. (1989). Underlying assumptions in the estimation of secondary structure content in proteins by circular dichroism spectroscopy – a critical review. *J. Pharm. Biomed. Anal.* 7, 1103-1119.

- Martin, S.R., and Bayley, P.M. (1986). The effects of  $\text{Ca}^{2+}$  and  $\text{Cd}^{2+}$  on the secondary and tertiary structure of bovine testis calmodulin. *Biochem. J.* 238, 485-490.
- Martin, S.R., and Bayley, P.M. (2002). Regulatory implications of a novel mode of interaction of calmodulin with a double IQ-motif target sequence from *murine* dilute myosin V. *Protein Sci.* 11, 2909-2923.
- Masino, L., Martin, S.R., and Bayley, P.M. (2000). Ligand binding and thermodynamic stability of a multidomain protein, calmodulin. *Protein Sci.* 9, 1519-1529.
- Mayhood, T.W., and Windsor, W.T. (2005). Ligand binding affinity determined by temperature-dependent circular dichroism: cyclin-dependent kinase 2 inhibitors. *Anal. Biochem.* 345, 187-197.
- Morillas M., Vanik D.L., and Surewicz, W.K. (2001). On the mechanism of alpha-helix to beta-sheet transition in the recombinant prion protein. *Biochemistry.* 40, 6982–6987.
- Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant  $m$  values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* 4, 2138-2148.
- Pace, C.N., and McGrath, T. (1980). Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.* 255, 3862-3865.
- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4, 2411-2423.
- Perczel, A., Hollósi, M., Tusnády, G., and Fasman, G.D. (1991). Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* 4, 669-679.
- Provencher, S.W., and Glöckner, J (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20, 33-37.
- Robertson, A.D., and Murphy, K. P. (1997). Protein structure and the energetics of protein stability. *Chem. Rev.* 97, 1251-1267.

- Saxena, V.P., and Wetlaufer, D.B. (1971). A new basis for interpreting the circular dichroism spectra of proteins. *Proc. Natl. Acad. Sci. USA* 68, 969-972.
- Schellman, J.A. (1975). Macromolecular binding. *Biopolymers* 14, 999-1018.
- Sreerama, N., and Woody, R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* 209, 32-44.
- Sreerama, N. and Woody, R. W. (1994). Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods. *J. Mol. Biol.* **242**, 497-507.
- Sreerama, N., and Woody, R.W. (2000a). Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* 287: 243-251.
- Sreerama, N., and Woody, R.W. (2000b). Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287: 252-260.
- Sreerama, N., and Woody, R.W. (2004a). Computation and analysis of protein circular dichroism spectra. *Meth. Enzymol.* 383, 318-351
- Sreerama, N., and Woody, R. W. (2004b). On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* 13, 100-112.
- Sreerama, N., Venyaminov, S.Y., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* 287, 243-251.
- Strickland, E. H. (1974). Aromatic contributions to circular dichroism spectra of proteins. *CRC Crit. Rev. Biochem.* 2, 113-175.

- Swietnicki, W., Morillas, M., Chen, S.G., Gambetti, P., and Surewicz, W.K. (2000). Aggregation and fibrillization of the recombinant human prion protein huPrP<sub>90-231</sub>. *Biochemistry* 39, 424–431.
- van Stokkum, I.H.M., Spoelder, H.J.W., Bloemendal, M., van Grondelle, R., and Groen, F.C.A. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.* 191, 110-118.
- Venyaminov, S. Y., and Yang, J. T. (1996). Determination of protein secondary structure. In “Circular dichroism and the conformational analysis of biomolecules” (G.D. Fasman, ed.), pp. 69-107. Plenum Press, New York.
- Wallace, B.A., Lees, J. G., Orry, A.J.W., Lobley, A., and Janes, R.W. (2003). Analyses of circular dichroism spectra of membrane proteins. *Protein Sci.* 12, 875-884.
- Woody, R. W. (1985). Circular dichroism of peptides. In “The Peptides” (V.J. Hruby, ed.), Volume 7, pp 15-114. Academic Press, New York.
- Woody, R. W. (1995). Circular dichroism, *Methods Enzymol.* **246**, 34-71.
- Woody, R. W. (1996). Theory of circular dichroism of proteins. In “Circular dichroism and the conformational analysis of biomolecules” (G.D. Fasman, ed.), pp. 25-67. Plenum Press, New York.
- Woody, R. W., and Dunker, A. K. (1996). Aromatic and cystine side-chain circular dichroism in proteins. In “Circular dichroism and the conformational analysis of biomolecules” (G.D. Fasman, ed.), pp. 109-157. Plenum Press, New York.
- Yang, J.T., Wu, C-S.C., and Martinez, H.M. (1986). Calculation of protein conformation from circular dichroism. *Meth. Enzymol.* 130, 208-269.

## Figure Legends

### Figure 1 – Typical Near-UV CD spectra

(A) Near-UV CD spectra of calcium-saturated (**a**) and calcium-free (**b**) *Drosophila* calmodulin. The nine phenylalanines in this protein produce the sharp bands at 262 and 268 nm; the single tyrosine in the C-terminal domain produces the broad band around 275 nm. Note that calcium binding reverses the sign of the tyrosine band.

(B) Near-UV CD spectra of calcium-saturated *Drosophila* calmodulin (**a**) and the target peptide from CaM-dependent kinase I (**b**). The spectrum of the 1:1 complex (**c**) is clearly not equal to the sum of the component spectra and the difference spectrum (**c-(a+b)**) shows the characteristic features of tryptophan CD (see text and Clapperton et al. (2002)).

### Figure 2 – Typical Far-UV CD spectra

Far-UV CD spectra of (**a**) myoglobin (All- $\alpha$ : 4mbn.pdb); (**b**) prealbumin (All- $\beta$ : 2pab.pdb), and (**c**) acid denatured staphylococcal nuclease at pH 6.2 and 6°C (Unordered). These spectra are available on the internet (at <http://lamar.colostate.edu/~sreeram/CDPro/>).

### Figure 3 – pH-induced conformational transitions.

The short peptide LKLKLLKLLKLLKLLKLG was designed to form a perfect amphipathic helix. However, at low pH it is unordered (curve **a**), probably because of charge repulsion between the lysines. Neutralization of the lysines at high pH permits formation of the  $\alpha$ -helix (curve **b**) and the unordered to  $\alpha$ -helical transition is readily monitored using CD (**inset**). (Schilstra, M.J., unpublished observations).

### Figure 4 – Binding of metal ions to calmodulin.

Binding of calcium (**a**) and terbium (**b**) to *Drosophila* calmodulin monitored using tyrosine CD (see Fig. 1A). Calmodulin is a two-domain protein with a total of four metal ion-binding

sites, two in each domain. Because this particular calmodulin contains a single tyrosine located in the C-terminal domain these stoichiometric titrations show that calcium binds preferentially to the C-terminal tyrosine-containing domain (>75% saturation at  $[Ca]/[CaM] = 2$ ), whilst terbium (a calcium analogue) binds preferentially to the N-terminal domain (< 25% saturation at  $[Tb]/[CaM] = 2$ ). (Martin, S.R., unpublished observations).

### **Figure 5 – Protein stability curves.**

Calculated variation of  $\Delta G^\circ$  as a function of temperature for three proteins with the same  $\Delta H^\circ_{T_m}$  (90 kcal/mol) but different values of  $T_m$  and  $\Delta C_p$ . **(a)**  $T_m = 60^\circ\text{C}$ ,  $\Delta C_p = 1.2$  kcal/mol.K; **(b)**  $T_m = 80^\circ\text{C}$ ,  $\Delta C_p = 1.6$  kcal/mol.K; **(c)**  $T_m = 80^\circ\text{C}$ ,  $\Delta C_p = 2.0$  kcal/mol.K

### **Figure 6 – Titration saturation curves.**

Observed fractional saturation for a titration of A with B in which a simple 1:1 complex (AB) is formed (see text). The fractional saturation of A (defined as  $f_A = [AB]/A_T$ ) is plotted as a function of the  $B_T/A_T$  ratio for a system with  $K_d = 1 \mu\text{M}$  and different values of  $A_T$ : **(a)**  $A_T = 2 \mu\text{M}$  ( $2 \times K_d$ ), **(b)**  $A_T = 25 \mu\text{M}$  ( $25 \times K_d$ ), and **(c)**  $A_T = 0.2 \mu\text{M}$  ( $0.2 \times K_d$ ). Note:  $A_T$  and  $B_T$  indicate total concentrations of A and B (see text for further details).

Figure 1

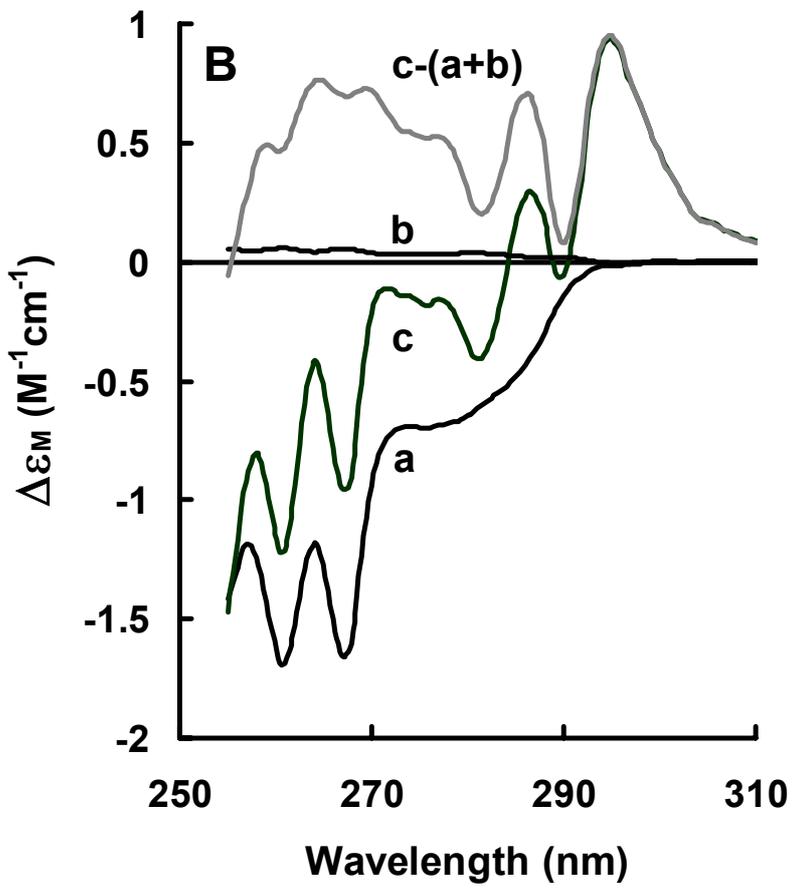
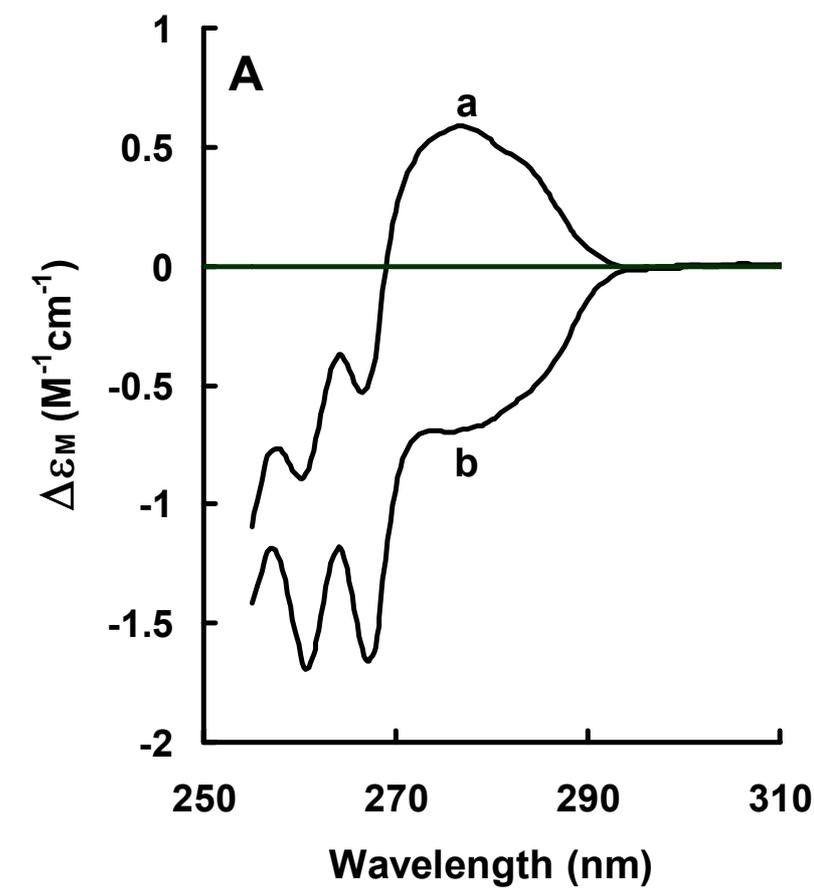


Figure 2

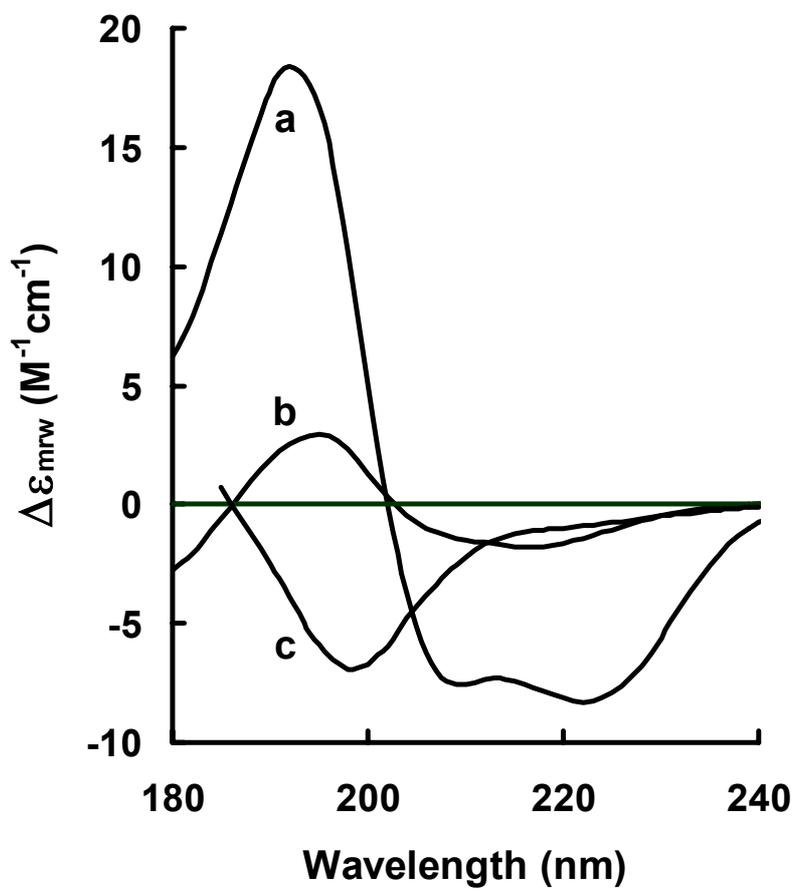


Figure 3

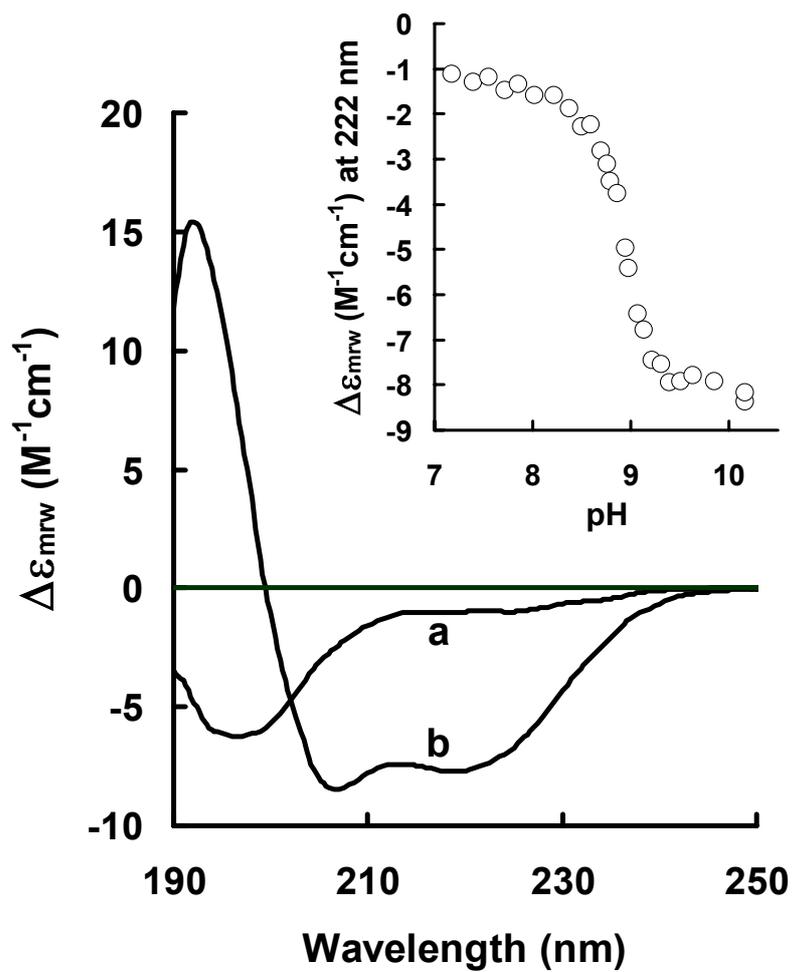


Figure 4

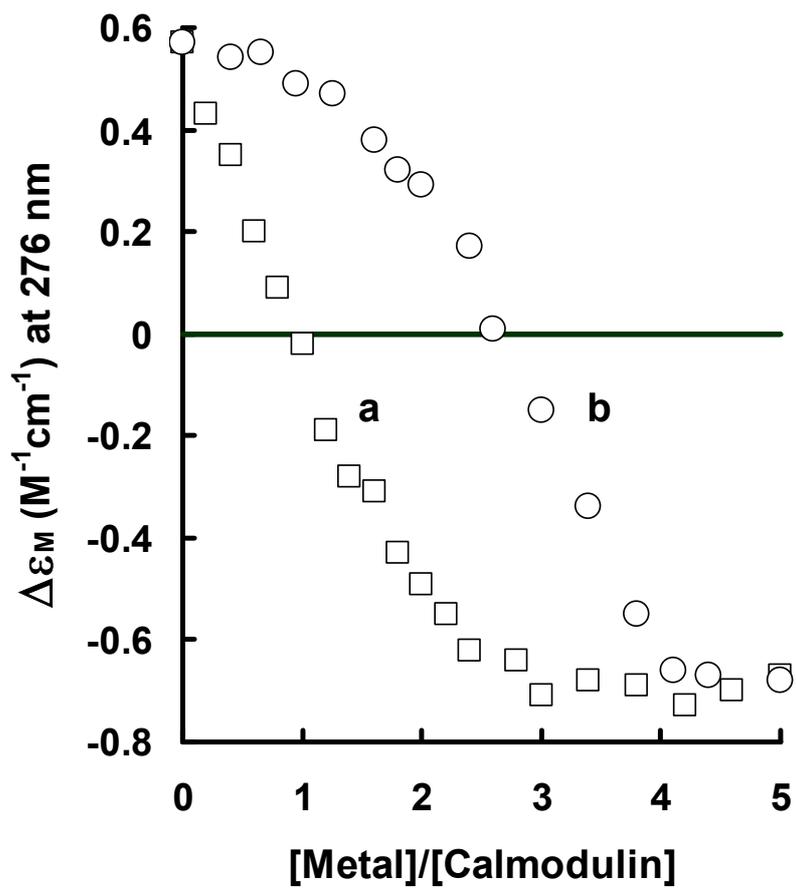


Figure 5

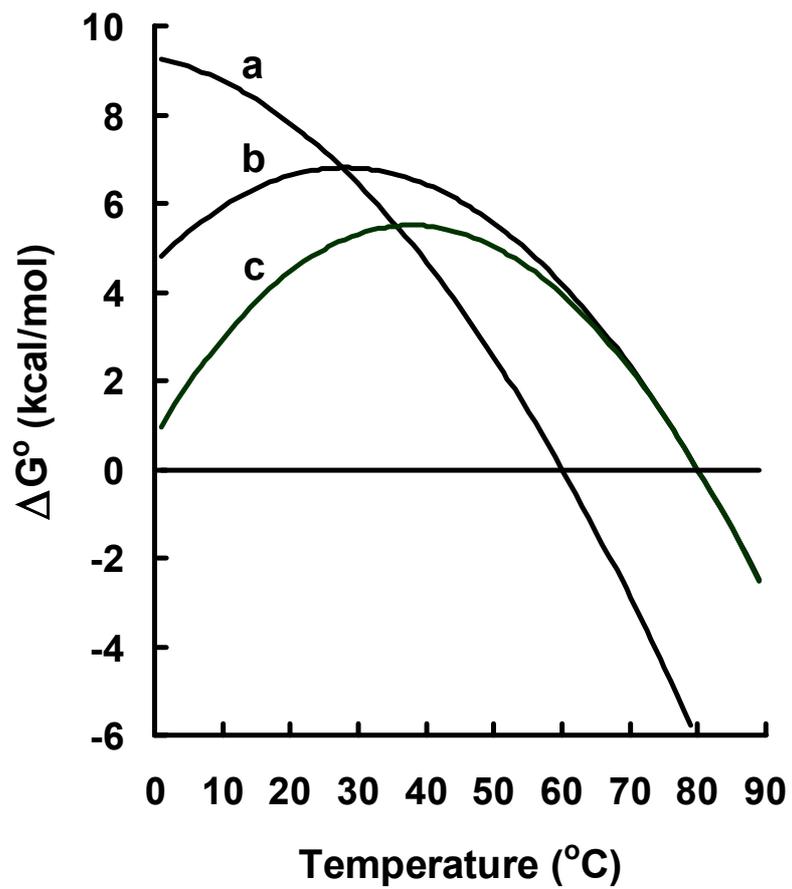


Figure 6

